AD-A104 476 JOHNS HOPKINS UNIV BALTIMORE MD DEPT OF MATHEMATICAL--ETC F/G 12/1
A NOTE ON A NONPARAMETRIC MAXIMUM PENALIZED LIKELIHOOD ESTIMATO--ETC(U)
APR 81 V K KLONIAS N00014-79-C-0801
UNCLASSIFIED TR-337 NL

1 OF 1
AD A
104476

END
DATE
FILMED
10-81
DTIC

ADA104476

DEPARTMENT OF MATHEMATICAL SCIENCES
The Johns Hopkins University
Baltimore, Maryland 21218

LEVEL

A NOTE ON A NONPARAMETRIC MAXIMUM PENALIZED LIKELIHOOD ESTIMATOR OF THE PROBABILITY DENSITY FUNCTION OF A POSITIVE RANDOM VARIABLE, A MPLE WITH POSITIVE SUPPORT

By

(10) V. K. Klonias
The Johns Hopkins University (12) 9

(9) Technical Report No. 337
ONR Technical Report No. 81-2
(11) April, 1981

~~(18) ONR~~

~~(19) TR-81-2~~

14 TR-337, TR-81-2

DTIC ELECTE SEP 23 1981 S D
D

(15)

Research supported in part by the Army, Navy and Air Force under Office of Naval Research Contract No. N00014-79-C-0801. Reproduction in whole or part is permitted for any purpose of the United States Government.

81 9 22 036

408328

Summary

The "first nonparametric maximum penalized likelihood density estimator of Good and Gaskins", corresponding to a penalty proportional to the Fisher information, is derived in the case that the density function has its support on the half-line. The computational feasibility as well as the consistency properties of the estimator are indicated.

A NOTE ON A NONPARAMETRIC MAXIMUM PENALIZED LIKELIHOOD ESTIMATOR OF THE PROBABILITY DENSITY FUNCTION OF A POSITIVE RANDOM VARIABLE†

A MPLE WITH POSITIVE SUPPORT

By

V. K. KLONIAS

The Johns Hopkins University

AMS 1970 subject classification. Primary 62G10; Secondary 62E10.

Key words and phrases. Nonparametric density estimator with positive support, maximum penalized likelihood method, Fisher information functional, exponential spline function.

†Research supported in part by the Office of Naval Research Contract No. N00014-79-C-0801.

Let $X_1, X_2, \ldots, X_n$ be independent observations from a distribution function F with density function f assumed to have finite Fisher information $I(f) \equiv \int (f')^2/f = 4\int (v')^2$, where $v \equiv f^{\frac{1}{2}}$. The maximum penalized likelihood method of density estimation (MPLE) was introduced by Good and Gaskins (1971) and consists of maximizing the penalized likelihood functional $L(f) \equiv \prod_{i=1}^{n} f(X_i) \exp\{-\Phi(f)\}$, where Φ denotes some penalty functional for "rough" density functions f. Thus, they avoided the Dirac delta solution of the unpenalized problem, and for the two penalty functionals they proposed they were led to two nonparametric density estimators, known as the "first and second MPLE's of Good and Gaskins" after de Montricher, Tapia and Thompson's (1975) paper where their existence and uniqueness were rigorously established within the framework of Sobolev spaces.

The "first MPLE of Good and Gaskins" f_n, to which we restrict ourselves here, corresponds to $\Phi(f) = \alpha I(f)/4$, $\alpha > 0$, and in the case that the support of f is the entire real line $\mathbb{R}$ and $v \in H^1(\mathbb{R}) \equiv \{v\colon v, v' \in L_2(\mathbb{R})\}$ - a Sobolev space of order one - de Montricher et al (1975) showed f_n to be an exponential spline with knots at the sample points, given by $f_n = u_n^2$, where

$$(1) \quad u_n(x) = (4\lambda_n\alpha)^{-\frac{1}{2}} \sum_{i=1}^{n} u_n(X_i)^{-1} \exp\{-(\lambda_n/\alpha)^{\frac{1}{2}} |x - X_i|\}, \; x \in \mathbb{R},$$

is the MPLE of v, with $\lambda_n > 0$ - the Lagrange multiplier corresponding to the constraint $\int f = 1$ of the underlying optimization problem.

We will show that in the case that f has its support on the half line $\mathbb{R}_+ \equiv (0,\infty)$ and $v \in H^1(\mathbb{R}_+)$, the "first MPLE of Good and Gaskins" f_+ (we supress the subscript n) is also an exponential spline with knots at the sample points, given by $f_+ = u_+^2$, where u_+ - the MPLE of v - is given by (2) below.

Let $\|\cdot\|_2$, $\|\cdot\|_{2,+}$ denote the $L_2(\mathbb{R})$ and $L_2(\mathbb{R}_+)$ norms respectively, and consider the MPLE problem

(P1) $\max \prod_{i=1}^{n} u^2(X_i)\exp\{-\alpha \|u'\|_{2,+}^2\}, \quad u \in H^1(\mathbb{R}_+)$

subject to: $\|u\|_{2,+} = 1$ and $u(X_i) \geq 0$, $i = 1,2,\ldots,n$.

<u>Proposition 1</u>. Problem (P1) has a unique solution u_+, given implicitly by

(2) $$u_+(x) = (4\lambda\alpha)^{-\frac{1}{2}} \sum_{i=1}^{n} u_+(X_i)^{-1}[\exp\{-(\lambda/\alpha)^{\frac{1}{2}} |x-X_i|\} + \exp\{-(\lambda/\alpha)^{\frac{1}{2}}|x+X_i|\}], \quad x \in \mathbb{R}_+,$$

where $\lambda > 0$ is the Lagrange multiplier corresponding to the constraint $\|u\|_{2,+} = 1$.

<u>Proof</u>. Let $\bar{u}(x) \equiv u(|x|)$ for all $x \in \mathbb{R}\backslash\{0\}$, $\bar{u}(0) \equiv \lim_{x\to 0^+} u(x)$, and set $X_{-i} \equiv X_i$ for all $i=1,\ldots,n$. Then problem (P1) is equivalent to problem

(P2) $\max \prod_{|i|=1}^{n} \bar{u}^2(X_i) \exp\{-\alpha\|\bar{u}'\|_2^2\}, \quad \bar{u} \in H_s$

subject to: $\|\bar{u}\|_2^2 = 2$ and $\bar{u}(X_i) \geq 0$, $|i| = 1,\ldots,n$,

where $H_s \equiv \{g \in H^1(\mathbb{R}) : g(x) = g(-x)$ for all $x \in \mathbb{R}\}$. Notice that for $\bar{u} \in H^1(\mathbb{R})$, i.e., for $\bar{u}$ not necessarily symmetric, there exists a unique solution to problem (P2) given by

$$\bar{u}_0(x) = (4\lambda\,\alpha)^{-\frac{1}{2}} \sum_{|i|=1}^{n} \bar{u}_0(X_i)^{-1} \exp -(\lambda/\alpha)^{\frac{1}{2}} |x-X_i|\}, \; x \in \mathbb{R},$$

where λ is the Lagrange multiplier corresponding to the constraint $\|\bar{u}\|_2^2 = 2$. The arguments leading to this result are identical to those in de Montricher et al (1975) leading to (1). Hence to show that the spline function $\bar{u}_0$ is also the unique solution to problem (P2) and hence $u_+(x) \equiv \bar{u}(x)$ for $x \in \mathbb{R}_+$, the unique solution to problem (P1), we need only prove that $\bar{u}$ is in H_s - i.e., symmetric about zero. To this end notice that $\bar{u}_0$ is smmetric everywhere if it is symmetric at the knots, i.e., if $\bar{u}(X_i) = \bar{u}(-X_i)$ for $i=1,\ldots,n$. But this is true since in system (3) below the variables $\bar{u}(X_i)$, $\bar{u}(-X_i)$, $i=1,\ldots,n$ are interchangeable:

$$(3) \quad \bar{u}(X_j) = (4\lambda\,\alpha)^{-\frac{1}{2}} \sum_{i=1}^{n} [\bar{u}(X_i)^{-1} \exp\{-(\lambda/\alpha)^{\frac{1}{2}} |X_j-X_i|\} + \bar{u}(-X_i)^{-1} \exp\{-(\lambda/\alpha)^{\frac{1}{2}} |X_j+X_i|\}],$$

$$\bar{u}(-X_j) = (4\lambda\,\alpha)^{-\frac{1}{2}} \sum_{i=1}^{n} [\bar{u}(X_i)^{-1} \exp\{-(\lambda/\alpha)^{\frac{1}{2}} |X_j+X_i|\} + \bar{u}(-X_i)^{-1} \exp\{-(\lambda/\alpha)^{\frac{1}{2}} |X_j-X_i|\}],$$

$j=1,\ldots,n$.

<u>Corollary 1</u>. The "first MPLE of Good and Gaskins" when f has its support on $\mathbb{R}_+$ is given by $f_+ = u_+^2$.

Proof. This is a consequence of the nonnegativity of u_+ and Lemma 3.1 in de Montricher et al (1975).

Remark 1. All the consistency results developed in Klonias (1981) for $f_n = u_n^2$, where u_n is given by (1), are also valid for f_+ and very little has to be changed in the way of proofs.

Remark 2. Equation (2) gives u_+ only implicitly and the values of the estimate at the sample points have to be determined, i.e., system (3) has to be solved and λ to be chosen so that $\|\bar{u}\|_2^2 = 2$. In Chapter 4 of Klonias (1980), utilizing the particular structure of the "first MPLE of Good and Gaskins", an efficient method is presented for the resolution of the spline f_n, which can be easily adapted to determine the values of f_+ at the knots. The reader is also referred to Good and Gaskins (1971, 1980), Scott, Tapia and Thompson (1976), Tapia and Thompson (1978), and Ghorai and Rubin (1979), where methods for the numerical evaluation of f_n are presented.

REFERENCES

[1] De Montricher, G. F., Tapia, R. A. and Thompson, J. R., (1975). Nonparametric maximum likelihood estimation of probability densities by penalty function methods. Ann. Statist., 3, 1329-1348.

[2] Ghorai, J. and Rubin, H. (1979). Computational procedure for maximum penalized likelihood estimate. J. Statist. Comput. Simul., 10, 65-78.

[3] Good, I. J. and Gaskins, R. A. (1971). Nonparametric roughness penalties for probability densities. Biometrika, 58, 2, 255-277.

[4] Good, I. J. and Gaskins, R. A. (1980). Density estimation and bumphunting by the penalized likelihood method exemplified by scattering and meteorite data. (Invited paper), J. Amer. Statist. Assoc., 75, 42-73.

[5] Klonias, V. K. (1980). Nonparametric density estimation: Contributions to the maximum penalized likelihood method. Ph.D. Dissertation, The University of Rochester, Rochester, N.Y.

[6] Klonias, V. K. (1981). Consistency of a nonparametric maximum penalized likelihood estimator of the probability density function. Technical Report No. 334, Department of Mathematical Sciences, The Johns Hopkins University.

[7] Scott, D. W., Tapia, R. A. and Thompson, J. R. (1976). Computer Science and Statistics: Ninth Annual Symposium on the Interface.

[8] Tapia, R. A. and Thompson, J. R. (1978). Nonparametric Probability Density Estimation. The Johns Hopkins University Press, Baltimore and London.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER ONR No. 81-2	2. GOVT ACCESSION NO. AD-A104 476
3. RECIPIENT CATALOG NUMBER	
4. TITLE A NOTE ON A NONPARAMETRIC MAXIMUM PENALIZED LIKELIHOOD ESTIMATOR OF THE PROBABILITY DENSITY FUNCTION OF A POSITIVE RANDOM VARIABLE A HOLE WITH POSITIVE SUPPORT	5. TYPE OF REPORT & PERIOD COVERED Technical Report
	6. PERFORMING ORGANIZATION REPORT NO. Technical Report No. 337
7. AUTHOR(s) V. K. Klonias	8. CONTRACT OR GRANT NUMBER(s) ONR No. N00014-79-C-0801
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Mathematical Sciences The Johns Hopkins University Baltimore, MD 21218	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME & ADDRESS Office of Naval Research Statistics & Probability Program Arlington, VA 22217	12. REPORT DATE April 1981
	13. NUMBER OF PAGES 7
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Nonparametric density estimator with positive support, maximum penalized likelihood method, Fisher information functional, exponential spline function.

20. ABSTRACT

The "first nonparametric maximum penalized likelihood density estimator of Good and Gaskins", corresponding to a penalty proportional to the Fisher information, is derived in the case that the density function has its support on the half-line. The computational feasibility as well as the consistency properties of the estimator are indicated.